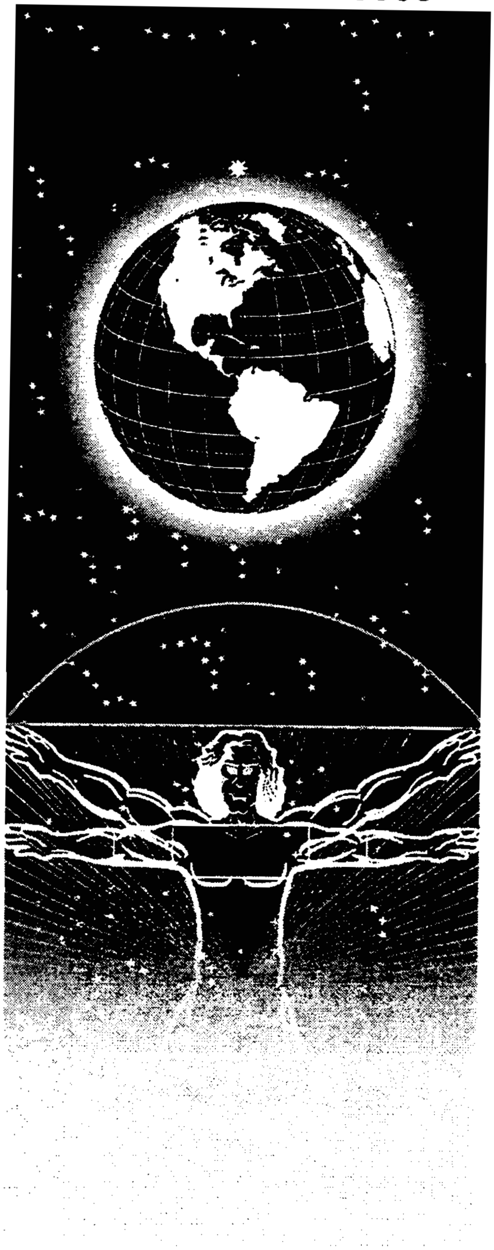
AL/HR-TP-1996-0050



UNITED STATES AIR FORCE
ARMSTRONG LABORATORY

Using Observer Ratings to Assess Situational Awareness in Tactical Air Environments

Herbert H. Bell

Wayne L. Waag

March 1997

Approved for public release; distribution is unlimited.

Human Resources Directorate
Aircrew Training Research Division
6001 South Power Road, Bldg 558
Mesa Arizona 85206-0904

NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

HERBERT H. BELL
Project Scientist

ELIZABETH L. MARTIN
Technical Director

LYNN A. CARROLL, Colonel, USAF
Chief, Aircrew Training Research Division

Please notify AL/HRPP, 7909 Lindbergh Drive, Brooks AFB, TX 78235-5352, if your address changes, or if you no longer want to receive our technical reports. You may write or call the STINFO Office at DSN 240-3877 or commercial (210) 536-3877.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 1997	Final Report - January 1994 to November 1995

4. TITLE AND SUBTITLE

Using Observer Ratings to Assess Situational Awareness in Tactical Air Environments

5. FUNDING NUMBERS

PE - 62202F
PR - 1123
TA - B3
WU - 02

6. AUTHOR(S)

Herbert H. Bell
Wayne L. Waag

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Armstrong Laboratory
Human Resources Directorate
Aircrew Training Research Division
6001 South Power Road, Building 558
Mesa, AZ 85206-0904

8. PERFORMING ORGANIZATION REPORT NUMBER

AL/HR-TP-1996-0050

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Armstrong Laboratory
Human Resources Directorate
7909 Lindbergh Drive
Brooks Air Force Base, Texas 78235-5352

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

Armstrong Laboratory Technical Monitor: Dr Herbert H. Bell, (602) 988-6561.
This paper documents a presentation at the International Conference on Experimental Analysis and Measurement of Situation Awareness held 1-3 November 1995 in Daytona Beach FL.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This paper summarizes the results of initial attempts to measure situation awareness (SA) in operational fighter squadrons and in multiship air combat simulations. It also discusses the general problem of using subjective measures to assess performance. Initial efforts focused on three issues: (a) the definition of SA, (b) the degree to which pilots can reliably judge their fellow pilots in terms of SA, and (c) whether or not there is a relationship between such judgements and mission performance. This paper was also published in the conference proceedings.

14. SUBJECT TERMS

Air Combat; Air combat simulation; Flight simulation; Flight training; Observer ratings; Performance; Performance measurement; SA; Simulation; Situation awareness; Subjective measures; Tactical air environments; Training; Training Evaluation

15. NUMBER OF PAGES

14

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION ABSTRACT
Unclassified	Unclassified	Unclassified	UL

PREFACE

This paper documents work performed on situational awareness in tactical air environments and which was presented at the international conference on Experimental Analysis and Measurement of Situation Awareness which was held at Daytona Beach, FL, from 1-3 November 1995. The presentation was also published in the conference proceedings. This paper summarizes initial attempts to measure situation awareness in operational fighter squadrons and in multiship air combat simualtions.

The effort was conducted under Work Unit 1123-B3-02, Tools for Assessing Situational Awareness. The principal investigator was Dr Herbert H. Bell.

Using Observer Ratings to Assess Situational Awareness in Tactical Air Environments

Herbert H. Bell and Wayne L. Waag

Armstrong Laboratory
Aircrew Training Research Division
6001 S. Power Road - Bldg. 558
Mesa, AZ 85206-0904

Introduction

In 1991, the Air Force Chief of Staff asked a series of questions about situational awareness (SA). These questions included: What is SA? Can we measure SA? Can we select individuals for pilot training based on their SA potential? What impact does training have on SA? In response to these questions, Armstrong Laboratory initiated an SA research program. This paper summarizes our initial attempts to measure SA in operational fighter squadrons and in multiship air combat simulations. It then discusses the general problem of using subjective measures to assess performance.

Our initial efforts have focused on three issues. The first issue concerns the definition of SA. The second issue is the degree to which pilots can reliably judge their fellow pilots in terms of SA. The third issue is whether or not there is a relationship between such judgments and mission performance.

In response to the question, "What is SA?," the Air Staff provided a working definition that links SA to mission performance. This definition, written from the operator's perspective, defines SA as "A pilot's continuous perception of self and aircraft in relation to the dynamic environment of flight, threats, and mission, and the ability to forecast, then execute tasks based on that perception (Carroll, 1992)." Although there are a number of other definitions of SA available (e.g., Endsley, 1995b; Rogers, 1992; Sarter & Woods, 1991; Tenney, Adams, Pew, Huggins), we are using this Air Staff definition as the basis for our research efforts. This definition reflects the importance of SA in mission accomplishment thus capturing the richness and complexity of the pilot's world. It emphasizes perceiving what is important and then using that perception to guide the selection and performance of appropriate behaviors. Unfortunately, it is also very complex because it combines processes, tasks, and the linkages between them into a single construct. Consequently, it is very difficult to separate SA from the other aspects of skilled performance that determine combat proficiency.

Measuring SA in Operational Fighter Squadrons

In order to determine whether or not pilots could reliably classify fellow pilots based upon SA, we limited our investigation to mission-ready F-15C pilots. With the assistance of instructor pilots and other subject-matter experts (SMEs), we developed a list of 31 behavioral elements of SA. Our SMEs felt these elements reflected SA and were important to mission success. Table 1 lists these 31 elements and the eight categories of mission performance they represent.

Table 1. Elements of Situational Awareness

General Traits	Information Interpretation
– Discipline	– Interpreting VSD
– Decisiveness	– Interpreting RWR
– Tactical knowledge	– Ability to use AWACS/GCI
– Time-sharing ability	– Integrating overall information
– Reasoning ability	– Radar sorting
– Spatial ability	– Analyzing engagement geometry
– Flight management	– Treat prioritization
Tactical Game Plan	System Operation
– Developing plan	– Radar
– Executing plan	– TEWS
– Adjusting plan on-the-fly	– Overall weapons system proficiency
Communication	Tactical Employment-BVR
– Quality (brevity, accuracy, timeliness)	– Targeting decisions
– Ability to effectively use information	– Fire-point selection
Tactical Employment-General	Tactical Employment-WVR
– Assessing offensiveness/defensiveness	– Maintain track of bogeys/friendlies
– Lookout (VSD, RWR, visual)	– Threat evaluation
– Defensive reaction (chaff, flares, maneuvering)	– Weapons employment
– Mutual support	

SA Instruments

The laboratory developed four different instruments to measure SA in operational F-15C squadrons based on the 31 elements listed in Table 1. The first instrument required respondents to provide their personal definition of SA. Using their personal definition of SA, each respondent then rated the importance of the 31 elements using a 6-point Likert scale.

The other three instruments, or SA Rating Scales (SARS), measured SA from three different perspectives: self, supervisory, and peer. All sample respondents completed the self-report and peer SARS. The self-report SARS and supervisory SARS required the respondents to rate either themselves or their subordinates on each of the 31 items. Both SARS used a 6-point scale and

the ratings were made relative to other F-15C pilots. The scale anchors were "Acceptable" and "Outstanding" because all respondents were on flying status and mission ready. The Squadron Commander, Operations Officer, Assistant Operations Officer, Weapons Officer, and Standardization-Evaluation Flight Examiner completed the supervisor SARS on the pilots within their squadron. In addition, squadron flight commanders completed supervisor SARS on the pilots within their flight. The peer SARS required respondents to rate the other mission-ready pilots in the squadron on general fighter pilot ability and SA ability and then to rank order them on their SA ability. Both the peer and supervisory SARS allowed respondents to omit rating a particular pilot if they felt they did not have enough information to accurately rate that individual.

Results

We obtained SA data from 238 mission-ready F-15 pilots from 11 squadrons stationed at four different Air Force bases. Two hundred and six of the respondents provided written definitions of SA. The first column in Table 2 lists the seven phases most frequently used by the respondents in defining SA. The second column shows the seven most highly rated elements of SA. There is considerable agreement between the phases used to define SA and the element ratings. In addition, both the phases and the element ratings indicate that a significant component of SA involves assimilating and using information to guide action.

Table 2. Phases Used to Define SA and Importance of SA Elements

Most Commonly Used Phases to Define SA	Most Highly Rated Elements for SA
– Composite 3-D image of entire situation	– Use of communication information
– Assimilation of information from multiple sources	– Information integration from multiple sources
– Knowledge of spatial position or geometric relationships among tactical entities	– Time-sharing ability
– Periodic mental update of dynamic situation	– Maintaining track of bogies and friendlies
– Prioritization of information and actions	– Adjusting plan on-the-fly
– Decision making quality	– Spatial ability to mentally picture engagement
– Projection of situation in time	– Lookout for threats from visual, RWR, VSD

Analyses of the peer and supervisory SARS indicated that the pilots can reliably classify their fellow pilots in terms of SA. Internal consistency was computed for all 31 items on the supervisory SARS. The resulting measure, Cronbach's coefficient α, was 0.99. Inter-rater reliability was also estimated for the supervisor and peer SARS using an analysis of variance procedure (Guilford, 1954). For the supervisor SARS, these analyses indicated that the average reliability of each supervisor's ratings was 0.50 and the average reliability of the pooled supervisor ratings was 0.88. Similarly, the peer SARS showed an individual reliability of 0.60 and a combined reliability of 0.97. Additional detail concerning the analyses of the SARS data is available in Waag and Houck (1994).

As shown in Table 3, there was substantial agreement between supervisor and peer SARS. Table 3 also indicates that there is noticeably less agreement between the self-report SARS and the other SARS.

Table 3. SARS Intercorrelations (N = 238).

	1	2	3	4	5
1. Supervisor SARS	--				
2. Peer -- Fighter pilot ability	.89	--			
3. Peer -- SA ability	.91	.98	--		
4. Peer -- Rank order	.92	.91	.92	--	
5. Self-report SARS	.45	.56	.57	.49	--

Measuring SA in Simulated Air Combat Missions

Although the SARS data indicate fairly high reliability and consistency between raters, they are not empirically linked to pilot performance in air combat missions. In an attempt to determine the relation between SA and mission performance, a composite SA score scaled with a mean 100 and a standard deviation 20 was computed for each of the 238 respondents. Based on this composite score, a sample of 40 mission-ready flight leads was selected to fly a series of multiship air-to-air combat simulations. The selected pilots covered the range of SA scores obtained for flight leads. An additional 23 mission-ready pilots flew as wingmen during the experiment. During each week-long SA simulation, the pilots flew nine sorties with four engagements per sortie. Sorties increased in complexity throughout the week.

Scenario Design

Figure 1 illustrates a typical scenario. In this defensive counterair mission, the two F-15s are defending an airfield. The attackers consist of two bombers escorted by two fighters. The simulation begins with the enemy force 80 nautical miles (nm) away from the airfield. The enemy fighters are flying at 20,000 ft and the bombers are at 10,000 ft. There is a lateral separation of 10 nm between the fighters and the bombers. At 35 nm, the fighters maneuver rapidly and descend to 3500 ft. At 15 nm, the bombers perform a hard right turn and descend to 2500 ft. The purpose of these maneuvers is to momentarily break the F-15s' radar contact and to disrupt the F-15 pilots' ability to identify, target, or engage the enemy aircraft.

Scenarios such as these contain events that "trigger" specific goal-directed behaviors necessary for mission accomplishment. We believe that SA can be inferred based on the pilot's reaction to such trigger events. In essence, these trigger events serve as SA probes in a naturalistic environment.

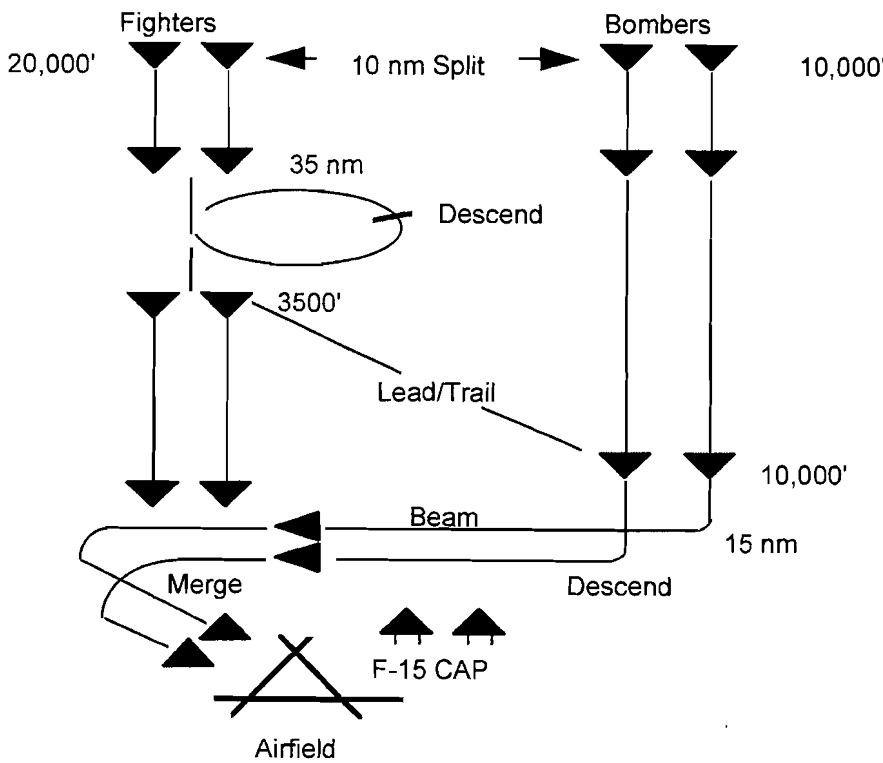


Figure 1. Defensive Counterair Mission Scenario

Rating Mission Performance

The basic approach taken toward SA measurement was through scenario manipulation and performance observation as suggested by Tenney, Adams, Pew, Huggins, and Rogers (1992). Other approaches, such as explicit probes and the Situation Awareness Global Assessment Technique (Endsley, 1995a), were considered. These other approaches were rejected because we needed measures that could be used during operational training either in simulators or actual aircraft.

As Kelly (1988) points out, measuring air combat skills presents a number of challenges. The fluid, dynamic nature of air combat, combined with the number of alternative tactics and techniques available to the pilot, make objective performance measurement extremely difficult. Even when objective data is available, it is often difficult to interpret the significance of that data. Because of the difficulties involved in interpreting air combat data, our approach is based on behavioral observation by SMEs who are unaware of the SA scores of the pilots they were observing. Two SMEs, retired fighter pilots with extensive experience in air combat and training, watched each engagement in real time and independently completed an observational checklist. To assist them in evaluating pilot performance, cockpit instruments, intraflight communications, and a plan view display of the engagement were available throughout the engagement. After each simulator session, the two SMEs discussed each engagement and completed a consensus performance rating scale containing 24 behavioral indicators based on the

SARS. In addition, the SMEs also wrote a critical event analysis for each mission that identified events that were critical to the outcome of the mission and indicative of the pilot's SA.

Results

Figure 2 shows the relationship between the composite SA scores obtained from the SARS and the mean SA score assigned by the SMEs based on their observation of performance during simulated air combat. The Pearson product moment correlation between these scores is 0.56. These data indicate that there is a significant relationship between squadron ratings of SA and performance in simulated air combat missions.

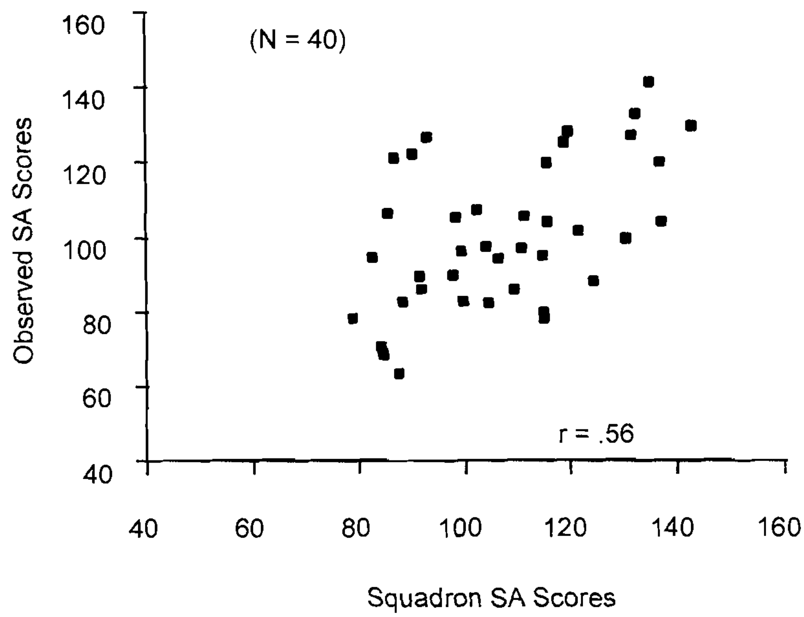


Figure 2. Simulator SA Scores and Squadron SA Scores

Discussion

We are encouraged by our initial results in developing measures of SA that can be used in a squadron's operational training environment. These results indicate that SA is a construct that has meaning and can be used by both peers and supervisors to classify mission-ready pilots. They also indicate that squadron ratings of SA are correlated with mission success in simulated air combat missions.

Although our approach to measurement may be classified as subjective rather than objective, we believe this is an oversimplification. All measurement approaches ultimately involve assigning numbers to events according to an explicit set of rules (Stevens, 1951). The distinction between objective and subjective measures simply indicates whether or not a human observer is an integral component of the measurement instrument. Objective measurement involves datum

that is generated independently of the human observer. Ideally, this datum is generated, recorded, and scored without the intervention of a human observer. Subjective measurement on the other hand, requires human observers to generate the datum itself. Although Muckler (1977) argues that there is no such thing as objective measurement in the strict sense, the distinction continues to be made and "so-called" objective measures are often preferred to subjective measures. The reason for this preference is that subjective measures are frequently seen as being contaminated by the human observers during the act of measurement. Since objective measures, on the other hand, are relatively independent of human observers, they are seen as "truer" measures of the construct under study.

Unfortunately, objective measures often fail to capture the richness and complexity of human performance (Kelly, 1988; Meister, 1989; Vreuls & Obermayer, 1985). One reason for this is that objective measures are essentially reductionistic and are therefore best suited for recording the fundamental dimensions of performance (e.g., latency, amount, and deviation). While these fundamental measures provide us with data that is less subject to error, they also frequently fail to provide us with information concerning the contextual nature of skilled performance. Subjective measures, on the other hand, seem more closely related to higher order psychological constructs. The datum they produce appears to reflect a synthesis of the more molecular behaviors and to reflect more global dimensions such as interpreting, judging, and deciding--the very essence of SA.

Obviously both measurement approaches are necessary if we are to develop our understanding of SA. The critical measurement issues are how do we refine our definition of SA and our measurement approaches and which measurements provide the best information for designing and evaluating aircrew training.

References

Carroll, L. A. (1992). Desperately seeking SA. *TAC Attack (TAC SP 127-1)*, 32, pp 5-6.

Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human Factors*, 37, pp 65-84.

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, pp 32-64.

Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw Hill.

Kelly, M. J. (1988). Performance measurement during simulated air-to-air combat. *Human Factors*, 30, pp 495-506.

Meister, D. (1989). *Conceptual aspects of human factors*. Baltimore, MD: Johns Hopkins University Press.

Muckler, F. A. (1977). Selecting performance measures: "Objective" versus "subjective" measurement. In L. T. Pope and D. Meister (Eds.), *Symposium proceedings: Productivity enhancement: Personnel performance assessment in Navy systems*. San Diego, CA: Navy Personnel Research and Development Center.

Sarter, N.B., and Woods, D.D. (1991). Situation awareness: A critical but ill-defined phenomenon. *International Journal of Aviation Psychology*, 1, pp 45-57.

Stevens, S. (1951). Mathematics, measurement, and psychophysics. In S. Stevens (Ed.), *Handbook of Experimental Psychology*. New York: Wiley.

Tenney, Y. J., Adams, J. J., Pew, R. W., Huggins, A. W. F., and Rogers, W. H. (1992). *A principled approach to the measurement of situation awareness in commercial aviation.* National Aeronautics and Space Administration, NASA Contractor Report 4451.

Vreuls, D., and Obermayer, R. W. (1985). Human-system performance measurement in training simulators. *Human Factors*, 27, pp 241-250.

Waag, W. L., and Bell, H. H. (1994). A study of situation assessment and decision making in skilled fighter pilots. Paper presented at the Second Conference on Naturalistic Decision Making, Dayton, OH, June.

Waag, W. L., and Houck, M. R. (1994). Tools for assessing situational awareness in an operational fighter environment. *Aviation, Space, and Environmental Medicine*, 65 (5, Suppl.), pp A13-A19.